CLAIMS

1. A method of indexing multimedia documents, the method being characterized in that it comprises the following steps:

a) for each document, identifying and extracting terms t_i constituted by vectors characterizing properties of the multimedia document for indexing, such as shape, texture, color, or structure of an image, the energy, the oscillation rate or frequency information of an audio signal, or a group of characters of a text;

b) storing the terms t_i characterizing the properties of the multimedia document in a term base comprising P terms;

c) determining a maximum number N of desired concepts combining the most pertinent terms t_i, where N is an integer less than P, with each concept c_i being designed to combine all terms that are neighboring from the point of view of their characteristics;

d) calculating the matrix T of distances between the terms t_i of the term base;

e) decomposing the set P of terms t_i of the term base into N portions P_j ($1 \le j \le N$) such that $P = P_1 \cup P_2 \ldots \cup P_j \ldots \cup P_N$, each portion P_j comprising a set of terms t_{ij} and being represented by a concept c_j, the terms t_i being distributed in such a manner that terms that are farther away are to be found in distinct portions P_l, P_m while terms that are closer together are to be found in the same portion P_l;

f) structuring a concept dictionary so as to constitute a binary tree in which the leaves contain the concepts c_i of the dictionary and the nodes of the tree contain the information necessary for scanning the tree during a stage of identifying a document by comparing it with previously-indexed documents; and

g) constructing a fingerprint base made up of the set of concepts c_i representing the terms t_i of the documents to be indexed, each document being associated with a fingerprint that is specific thereto.

2. An indexing method according to claim 1, characterized in that each concept c_i of the fingerprint base is associated with a data set comprising the number of terms No. T in the documents in which the concept c_i is present.

3. An indexing method according to claim 1 or claim 2, characterized in that for each document in which a concept c_i is present, a fingerprint of the concept c_i is registered in the document, said fingerprint containing the frequency with which the concept c_i occurs, the identities of concepts neighboring the concept c_i in the document, and a score which is a mean value of similarity measurements between the concept c_i and the terms t_i of the document that are the closest to the concept c_i.

4. An indexing method according to any one of claims 1 to 3, characterized in that it comprises a step of optimizing the partitioning of the set P of terms of the term base to decompose said set P into M classes C_i ($1 \leq i \leq M$, where $M \leq P$), so as to reduce the distribution error of the set P of terms in the term base into N portions ($P_1, P_2, \ldots, P_N$) where each portion P_i is represented by the term t_i that is taken as the concept c_i, the error that is committed ε being such that $\varepsilon = \sum_{i=1}^{N} \varepsilon_{t_i}$ where $\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$ is the error committed by replacing the terms t_j of a portion P_i with t_i.

5. An indexing method according to claim 4, characterized in that it comprises the following steps:

i) decomposing the set P of terms into two portions P_1 and P_2;

ii) determining the two terms t_i and t_j of the set P that are the furthest apart, corresponding to the greatest distance D_{ij} of the distance matrix T;

iii) for each term t_k of the set P, examining to see whether the distance D_{ki} between the term t_k of the term t_i is less that the distance D_{kj} between the term t_k and the term t_j, and if so, allocating the term t_k to the portion P_1, and otherwise allocating the term t_k to the portion P_2; and

iv) iterating step i) until the desired number N of portions P_i has been obtained, and on each iteration applying the steps ii) and iii) on the terms of the portions P_1 and P_2.

6. An indexing method according to claim 4 or claim 5, characterized in that it includes optimization starting from N disjoint portions $\{P_1, P_2, \ldots, P_N\}$ of the set P and N terms $\{t_1, t_2, \ldots, t_N\}$ representing them in order to reduce the decomposition error of the set P into N portions, and in that it comprises the following steps:

i) calculating the centers of gravity C_i of the portions P_i;

ii) calculating errors $\varepsilon C_i = \sum_{tj\in Pi} d^2(C_i, t_j)$ and $\varepsilon t_i = \sum_{tj\in Pi} d^2(t_i, t_j)$ when replacing the terms t_j of the portion P_i respectively by C_i and by t_i;

iii) comparing εt_i and εc_i and replacing t_i by C_i if $\varepsilon c_i \leq \varepsilon t_i$; and

iv) calculating a new distance matrix T between the terms t_i of the term base and the process of decomposing the set P of terms of the term base into N portions, unless a stop condition is satisfied with

$$\frac{\varepsilon c_t - \varepsilon c_{t+1}}{\varepsilon c_t} < \text{threshold},$$

where εc_t represents the error committed at instant t.

7. An indexing method according to any one of claims 1 to 6, characterized in that for the purpose of structuring the concept dictionary, a navigation chart is produced iteratively on each iteration, beginning by splitting the set of concepts into two subsets, and then selecting one subset on each iteration until the desired number of groups is obtained or until a stop criterion is satisfied.

8. An indexing method according to claim 7, characterized in that the stop criterion is constituted by the fact that the subsets obtained are all homogeneous with small standard deviation.

9. An indexing method according to claim 7 or claim 8, characterized in that during the structuring of the concept dictionary, navigation indicators are determined from a matrix $M = [c_1, c_2, \ldots, c_N] \in \Re^{p*N}$ of the set C of concepts $c_i \in \Re^p$ where c_i represents a concept of p values, by implementing the following steps:

i) calculating a representative w of the matrix M;

ii) calculating the covariance matrix $\tilde{M}$ between the elements of the matrix M and the representative w of the matrix M;

iii) calculating a projection axis u for projecting the elements of the matrix M;

iv) calculating the value $pi = d(u,Ci) - d(u,w)$ and decomposing the set of concepts C into two subsets C1 and C2 as follows:

$$\begin{cases} c_i \in C1 \text{ if } pi \le 0 \\ c_i \in C2 \text{ if } pi > 0 \end{cases}$$

v) storing the information $\{u, w, |p1|, p2\}$ in the node associated with C, where p1 is the maximum of all $pi \le 0$ and p2 is the minimum of all $pi > 0$, the data set $\{u, w, |p1|, p2\}$ constituting the navigation indicators in the concept dictionary.

10. An indexing method according to any one of claims 1 to 9, characterized in that both the structural components and the complements of said structural components constituted by the textural components of an image of the document are analyzed, and in that:

a) while analyzing the structural components of the image:

a1) boundary zones of the image structures are distributed into different classes depending on the orientation of the local variation in intensity so as to define structural support elements (SSEs) of the image; and

a2) performing statistical analysis to construct terms constituted by vectors describing the local properties and the global properties of the structural support elements;

b) while analyzing the textural components of the image:

b1) detecting and performing parametric characterization of a purely random component of the image;

b2) detecting and performing parametric characterization of a periodic component of the image; and

b3) detecting and performing parametric characterization of a directional component of the image;

c) grouping the set of descriptive elements of the image in a limited number of concepts constituted firstly by the terms describing the local and global properties of structural support element and secondly by the parameters of the parametric characterizations of the random, periodic, and directional components defining the textural components of the image; and

d) for each document, defining a fingerprint from the occurrences, the positions, and the frequencies of said concepts.

11. An indexing method according to claim 10, characterized in that the local properties of the structural support elements taken into consideration for constructing terms

comprise at least the support types selected from amongst a linear strip or a curved arc, the length and width dimensions of the support, the main direction of the support, and the shape and the statistical properties of the pixels constituting the support.

12. An indexing method according to claim 10 or claim 11, characterized in that the global properties of the structural support element taken into account for constructing terms comprise at least the number of each type of support and the spatial disposition thereof.

13. An indexing method according to any one of claims 10 to 12, characterized in that during analysis of the structural components of the image, a prior test is performed to detect whether at least one structure is present in the image, and in the absence of any structure, the method passes directly to the step of analyzing the textural components of the image.

14. An indexing method according to any one of claims 10 to 13, characterized in that in order to decompose boundary zones of the image structures into different classes, starting from the digitized image defined by the set of pixels $y(i,j)$ where $(i,j) \in I \times J$, where I and J designate respectively the number of rows and the number of columns of the image, the vertical gradient image $g_v(i,j)$ where $(i,j) \in I \times J$ and the horizontal gradient image $g_h(i,j)$ with $(i,j) \in I \times J$ are calculated, and the image is partitioned depending on the local orientation of its gradient into a finite number of equidistant classes, the image containing the orientation of the gradient being defined by the equation:

$$O(i,j) = \arctan\left[\frac{gh(i,j)}{gv(i,j)}\right]$$

the classes constituting support regions likely to contain significant support elements are identified, and on the basis

of the support regions, significant support elements are determined and indexed using predetermined criteria.

15. An indexing method according to any one of claims 1 to 9, characterized in that while indexing a multimedia document comprising video signals, terms t_i are selected that are constituted by key-images representing groups of consecutive homogeneous images, and concepts c_i are determined by grouping together terms t_i.

16. An indexing method according to claim 15, characterized in that in order to determine key-images constituting terms t_i, a score vector SV is initially generated comprising a set of elements SV(i) representative of the difference or similarity between the content of an image of index i and the content of an image of index i-1, and the score vector SV is analyzed in order to determine key-images which correspond to maximums of the values of the elements SV(i) of the score vector SV.

17. An indexing method according to claim 16, characterized in that an image of index j is considered as being a key-image if the value SV(j) of the corresponding element of the score vector SV is a maximum and the value SV(j) is situated between two minimums minL and minR, and if the minimum M1 such that M1 = $(|VS_{(j)} - minL|, |SV_{(j)} - minR|)$ is greater than a given threshold.

18. An indexing method according to any one of claims 1 to 9, characterized in that while indexing a multimedia document comprising audio components, the document is sampled and decomposed into frames, which frames are subsequently grouped together into clips each being characterized by a term t_i constituted by a parameter vector.

19. An indexing method according to claim 18, characterized in that a frame comprises about 512 samples to about 2,048 samples of the sampled audio document.

20. An indexing method according to claim 18 or claim 19, characterized in that the parameters taken into account to define the terms t_i comprise time information corresponding to at least one of the following parameters: the energy of the audio signal frames, the standard deviation of frame energies in the clips, the sound variation ratio, the low energy ratio, the rate of oscillation about a predetermined value, the high rate of oscillation about a predetermined value, the difference between the number of oscillation rates above and below the mean oscillation rate for the frames of the clips, the variance of the oscillation rate, the ratio of silent frames.

21. An indexing method according to any one of claims 18 to 20, characterized in that the parameters taken into account for defining the terms t_i comprise frequency information corresponding to at least one of the following parameters: the center of gravity of the frequency spectrum of the short Fourier transform of the audio signal, the bandwidth of the audio signal, the ratio between the energy in a frequency band to the total energy in the entire frequency band of the sampled audio signal, the mean value of spectrum variation of two adjacent frames in a clip, the cutoff frequency of a clip.

22. An indexing method according to any one of claims 18 to 21, characterized in that the parameters taken into account for defining the terms t_i comprise at least energy modulation at 4 Hz.

23. An indexing method according to any one of claims 1 to 14, characterized in that the shapes of an image of a document are analyzed using the following steps:

a) performing multiresolution followed by decimation of the image;

b) defining the image in polar logarithmic space;

c) representing the query image or image portion by its Fourier transform H;

d) characterizing the Fourier transform H as follows:

d1) projecting H in a plurality of directions to obtain a set of vectors of dimension equal to the projection movement dimension; and

d2) calculating the statistical properties of each projection vector; and

e) representing the shape of the image by a term t_i constituted by values for the statistical properties of each projection vector.